Linear Models: Permutation Methods

BRIAN S. CADE

Volume 2, pp. 1049–1054

in

Encyclopedia of Statistics in Behavioral Science

ISBN-13: 978-0-470-86080-9
ISBN-10: 0-470-86080-4

Editors

Brian S. Everitt & David C. Howell

Linear Models: Permutation Methods

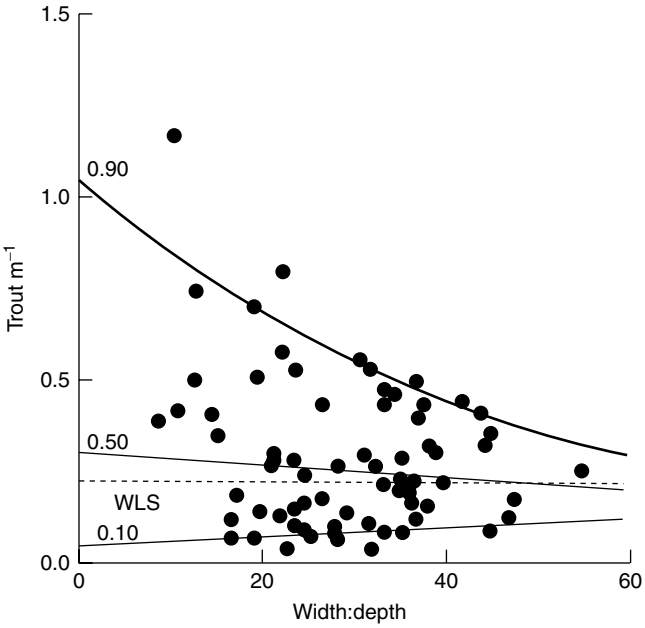
Permutation tests (*see* **Permutation Based Inference**) for the **linear model** have applications in behavioral studies when traditional parametric assumptions about the error term in a linear model are not tenable. Improved validity of Type I error rates can be achieved with properly constructed permutation tests. Perhaps more importantly, increased statistical **power**, improved robustness to effects of **outliers**, and detection of alternative distributional differences can be achieved by coupling permutation inference with alternative linear model estimators. For example, it is well-known that estimates of the mean in the linear model are extremely sensitive to even a single outlying value of the dependent variable compared to estimates of the median [7, 19]. Traditionally, linear modeling focused on estimating changes in the center of distributions (means or medians). However, quantile regression allows distributional changes to be estimated in all or any selected part of a distribution or responses, providing a more complete statistical picture that has relevance to many biological questions [6].

Parameters from the linear model in either its location, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, or location scale, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Gamma}\boldsymbol{\varepsilon}$, form can be tested with permutation arguments. Here, $\mathbf{y}$ is an $n \times 1$ vector of dependent responses, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters, $\mathbf{X}$ is an $n \times p$ matrix of predictors (with commonly the first column consisting of 1's for an intercept term), $\boldsymbol{\Gamma}$ is a diagonal $n \times n$ matrix where the n diagonal elements are the n corresponding ordered elements of the $n \times 1$ vector $\mathbf{X}\boldsymbol{\gamma}$ (diag($\mathbf{X}\boldsymbol{\gamma}$)), $\boldsymbol{\gamma}$ is a $p \times 1$ vector of unknown scale parameters, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors that are independent and identically distributed (iid) with density f_{ε}, distribution F_{ε}, and quantile F_{ε}^{-1} functions. Various parametric regression models are possible, depending on which parameter of the error distribution is restricted to equal zero; for example, setting the expected value $F_{\boldsymbol{\varepsilon}}(\boldsymbol{\mu}|\mathbf{X}) = 0$ yields the familiar mean regression, setting any quantile $F_{\boldsymbol{\varepsilon}}^{-1}(\boldsymbol{\tau}|\mathbf{X}) = 0$ yields quantile ($0 \leq \tau \leq 1$) regression, and the special case of $F_{\boldsymbol{\varepsilon}}^{-1}(0.5|\mathbf{X}) = 0$ yields median (least absolute deviation) regression [14]. The location model with homoscedastic error variance is just a special case of the linear-location scale model when $\boldsymbol{\gamma} = (1, 0, \ldots, 0)'$.

Estimates ($\hat{\boldsymbol{\beta}}$) of the various parametric linear models are obtained by minimizing appropriate loss functions of the residuals, $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. Minimizing the sum of squared **residuals** yields the **least squares estimates** of the mean model. Minimizing the sum of asymmetrically weighted (τ for $+$ residuals and $1 - \tau$ for $-$ and 0 residuals) absolute values of the residuals yields the quantile regression estimates, where least absolute deviation regression for the median ($\tau = 0.5$) model being just a special case [15]. Consistent estimates with reduced sampling variation can be obtained for linear location-scale models by implementing weighted versions of the estimators, where weights are the reciprocal of the scale parameters, $\mathbf{W} = \boldsymbol{\Gamma}^{-1}$. In applications, the $p \times 1$ vector of scale parameters $\boldsymbol{\gamma}$ would usually have to be estimated. Weighted regression estimates are obtained by multiplying $\mathbf{y}$ and $\mathbf{X}$ by $\mathbf{W}$ and minimizing the appropriate function of the residuals as before. Examples of various estimates are shown in Figure 1.

Test Statistic

A drop in dispersion, F-ratio-like, test statistic that is capable of testing hypotheses for individual or multiple coefficients can be evaluated by similar permutation arguments for any of the linear model estimators above [2, 5, 7, 19]. This pivotal test statistic takes the form $T = (S_{\text{reduced}} \div S_{\text{full}}) - 1$, where S_{reduced} is the sum minimized by the chosen estimator for the reduced parameter model specified by the null hypothesis ($\mathbf{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\xi}$) and S_{full} is the sum minimized for the full parameter model specified by the alternative hypothesis. This is equivalent to the usual F-ratio statistic for least squares regression but the degrees of freedom for reduced and full parameter models are deleted because they are not needed as they are invariant under the permutation arguments to follow. The reduced parameter model $\mathbf{y} - \mathbf{X}_2\boldsymbol{\xi} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$ is constructed by partitioning $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_1$ is $n \times (p - q)$ and $\mathbf{X}_2$ is $n \times q$; and by partitioning $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ where $\boldsymbol{\beta}_1$ is a $(p - q) \times 1$ vector of unknown nuisance parameters under the null and $\boldsymbol{\beta}_2$ is the $q \times 1$ vector of parameters specified by the null hypothesis $\mathbf{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\xi}$



τ	$\hat{\beta}_0$	90% CIβ_0	$\hat{\beta}_1$	90% CIβ_1	$P(H_0: \beta_1 = 0)$
0.90	0.058	–0.27, 0.40	–0.0215	–0.0321, –0.0125	0.001
0.50	–1.212	–1.90, –0.67	–0.0070	–0.0233, 0.0121	0.674
WLS	–1.475	–1.94, –1.01	–0.0010	–0.0142, 0.0123	0.900
0.10	–3.085	–3.92, –2.58	0.0163	0.0007, 0.0404	0.090

Figure 1 Lahontan cutthroat trout m^{-1} and width:depth ratios for small streams sampled 1993 to 1999 ($n = 71$) [10]; exponentiated estimates for 0.90, 0.50, and 0.10 regression quantiles and least squares (WLS) estimate for the weighted model $w(\ln y) = w(\beta_0 + \beta_1 X_1 + (\gamma_0 + \gamma_1 X_1)\varepsilon)$, $w = (1.310 - 0.0017X_1)^{-1}$. Tabled values are estimates, 90% confidence intervals from inverting the permutation test, and estimated probabilities for $H_0 : \beta_1 = 0$ based on the double permutation procedure for statistic T made with $m + 1 = 100000$ permutations

(frequently $\boldsymbol{\beta}_2 = \mathbf{0}$). The unconstrained, full parameter model is $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$. To test hypotheses on weighted estimates in the linear location-scale model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\Gamma\varepsilon}$, the terms $\mathbf{y}, \mathbf{X_1}$, and $\mathbf{X_2}$ are replaced with their weighted counterparts $\mathbf{Wy}, \mathbf{WX_1}$, and $\mathbf{WX_2}$, respectively, where $\mathbf{W}$ is the weights matrix, and the test statistic is constructed similarly. For homogeneous error models, $\mathbf{W} = \mathbf{I}$, where $\mathbf{I}$ is the $n \times n$ identity matrix.

When testing hypotheses on a single parameter, for example, $H_0 : \beta_j = \xi_j$, the more general T statistic may be replaced with an equivalent t-ratio form $t = (\hat{\beta}_j - \xi_j)/\sqrt{S_{\text{full}}}$, where $\hat{\beta}_j$ is the full model estimate of β_j [1, 2, 18]. Use of test statistics that are not pivotal such as the actual parameter estimates is not recommended because they fail to maintain valid Type I error rates when there is multicollinearity (*see* **Collinearity**) among the predictor variables in $\mathbf{X}$ [1, 2, 13].

Permutation Distribution of the Test Statistic

The test statistic for the observed data $T_{\text{obs}} = (S_{\text{reduced}} \div S_{\text{full}}) - 1$ is compared to a reference distribution of T formed by permutation arguments. The relevant exchangeable quantities under the null hypothesis to form a reference distribution are the errors $\boldsymbol{\varepsilon}$ from the null, reduced parameter model [1, 3, 17]. These can be permuted (shuffled) across the rows of $\mathbf{X}$ with equal probability. In general, the errors are unknown and unobservable. But, in the case of simultaneously testing all parameters other than the

intercept for the location model when $\mathbf{W} = \mathbf{I}$, $\mathbf{X_1}$ is an $n \times 1$ matrix of 1's, the residuals from the null, reduced parameter model $\mathbf{e_{red}} = \mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1$ or $\mathbf{y}$ differ from the errors $\boldsymbol{\varepsilon}$ by unknown constants that are invariant under permutation. Thus, in this case, a reference distribution for T can be constructed by either permuting $\mathbf{e_{red}}$ or $\mathbf{y}$ against the full model matrix $\mathbf{X}$ to yield probabilities under the null hypothesis (proportion of $T \geq T_{\text{obs}}$) that are exact, regardless of the unknown distribution of the errors [2, 5, 18]. The variables being tested in $\mathbf{X_2}$ may be continuous, indicators $(1, 0, -1)$ for categorical groups, or both. Thus, exact probabilities for the null hypothesis are possible for the 1-way classification (ANOVA) (*see* **Analysis of Variance**) model for two or more treatment groups, for the single-slope parameter in a simple regression, and for all slope parameters simultaneously in a **multiple regression**. In practice, for reasonable n, a very large random sample of size m is taken from the $n!$ possible permutations so that probability under the null hypothesis is estimated by (the number of $T \geq T_{\text{obs}} + 1)/(m + 1)$. The error of estimating the P value by Monte Carlo resampling (*see* **Monte Carlo Simulation**) can be made arbitrarily small by using a large m, for example, $m + 1 \geq 10000$.

For null hypotheses on subsets of parameters from a linear model with multiple predictors, the reference distribution of T is approximated by permuting $\mathbf{e_{red}}$ against $\mathbf{X}$ [1, 2, 5, 7, 11, 21]. Permuting against the full model matrix $\mathbf{X}$ ensures that the correlation structure of the predictor variables is fixed, that is, the design is ancillary. As the residuals no longer differ from the errors $\boldsymbol{\varepsilon}$ by a constant, they are not exchangeable with equal probability and the resulting probability for the null hypothesis is no longer exact [1, 5, 9, 12]. Regardless, this permutation approach originally due to Freedman and Lane [11] was found to have perfect correlation asymptotically with the exact test for least squares regression (as if the errors $\boldsymbol{\varepsilon}$ were known) [3] and has performed well in simulations for least squares [2], least absolute deviation [7], and quantile regression [5]. Some authors have permuted $\mathbf{e_{full}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ rather than $\mathbf{e_{red}}$ [1, 18, 21], but there is less theoretical justification for doing so, although it may yield similar results asymptotically for least squares regression estimates [2, 3]. There are alternative restricted permutation schemes that provide valid probabilities for linear models when the null hypothesis and hence $\mathbf{X}_2$ only include indicator variables for categorical treatment groups [1, 4, 19, 20]. Permutation tests for random (*see* **Completely Randomized Design**) and mixed effects in multifactorial linear models (*see* **Linear Multilevel Models**) are discussed in [4].

There are several approaches to improving exchangeability of the residuals $\mathbf{e_{red}}$ under the null hypothesis to provide more valid Type I error rates. For the linear least squares regression estimator, linear **transformations** toward approximate exchangeability of the residuals from a model with $p - q$ parameters must reduce the dimension of $\mathbf{e_{red}}$ to $n - (p - q)$ or $n - (p - q) + 1$ and reduce $\mathbf{X}$ (or $\mathbf{WX}$) to conform, for example, by Gram-Schmidt orthogonalization [9]. This theory is not directly applicable to a nonlinear estimator like that used for quantile regression. But reducing the dimension of $\mathbf{e_{red}}$ by deleting $(p - q) - 1$ of the zero residuals and randomly deleting $(p - q) - 1$ rows of $\mathbf{X}$ to conform was found to improve Type I error rates for null hypotheses involving subsets of parameters for both linear location and location-scale quantile regression models with multiple independent variables [8]. This approach was motivated by [9] and the fact that quantile regression estimates for $p - q$ parameters must have at least $p - q$ zero residuals. An example of the Type I error rates for corrected and uncorrected residuals $\mathbf{e_{red}}$ for a 0.90 quantile regression model is shown in Figure 2. Another option for quantile regression is to use a τ-rank score procedure on $\mathbf{e_{red}}$, which transforms the + residuals to τ, – residuals to $\tau - 1$, and zero residuals to values in the interval $(\tau - 1, \tau)$, which are then used to compute a test statistic similar to the one above but based on a least squares (or weighted least squares) regression of the rank scores $\mathbf{r}$ on $\mathbf{X}$ [5]. This τ-rank score test can be evaluated by permuting $\mathbf{r}$ across the rows of $\mathbf{X}$ to yield a reference distribution to compute a probability under the null hypothesis. An example of the Type I error rates associated with this procedure also is shown in Figure 2.

An additional complication with permutation testing for the linear models occurs whenever the null model specified by the hypothesis does not include an intercept term so that the estimates are constrained through the origin. This includes testing a null hypothesis that includes the intercept term or when testing subsets of weighted parameter estimates for variables that are part of the weights function. Residuals from the estimates for the null, reduced parameter model $\mathbf{e_{red}}$ are no longer guaranteed to be

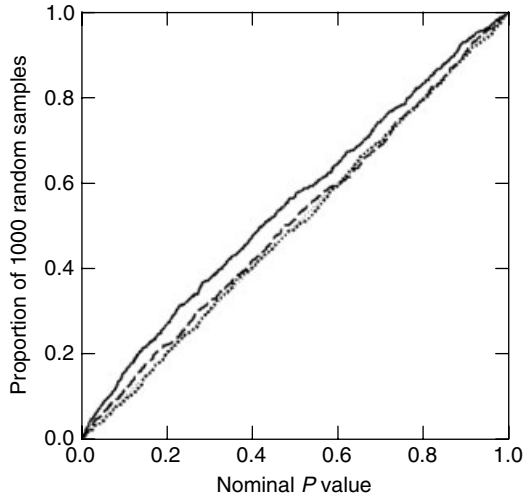


Figure 2 Cumulative distributions of 1000 estimated Type I errors for permutation tests of $H_0 : \beta_3 = \beta_5 = 0$ based on uncorrected permutation of raw residuals (solid line), deletion of $(p - q) - 1 = 3$ zero residuals (dashed line) prior to permutation, and by permuting τ-quantile rank scores (thick dotted line) for the 0.90 weighted quantile regression model $wy = w(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_3 X_4 + (1 + 0.05X_1)\varepsilon)$; $X_1 = \text{Uniform}(0, 100)$, $X_2 = 4000 - 20X_1 + N(0, 300)$, $X_3 = 10 + 0.4X_1 + N(0, 16)$, X_4 is 0,1 indicator variable with half of n randomly assigned each, $\beta_0 = 36$, $\beta_1 = 0.10$, $\beta_2 = -0.005$, $\beta_4 = 2.0$, $\beta_3 = \beta_5 = 0$, ε is lognormal $(F_{\varepsilon}^{-1}(0.9) = 0, \ \sigma = 0.75)$, $w = (1 + 0.05X_1)^{-1}$, and $n = 90$. Each P value was estimated with $m + 1 = 10000$ permutations. Fine dotted 1 : 1 line is the expected cdf

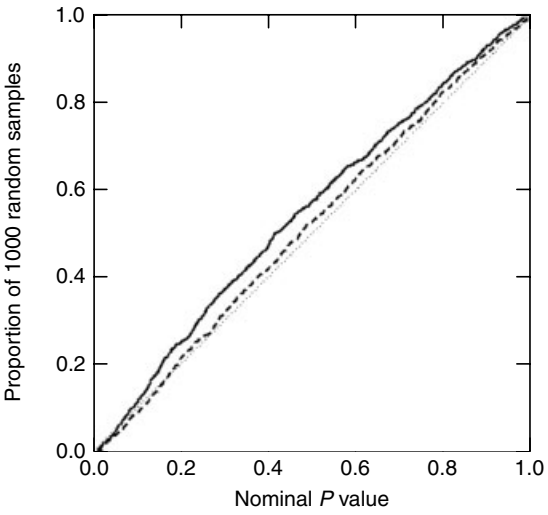


Figure 3 Cumulative distributions of 1000 estimated Type I errors for permutation tests of $H_0 : \beta_0 = 0$ based on uncorrected permutation of raw residuals (solid line) and double permutation of residuals (dashed line) for the 0.50 quantile regression model $y = \beta_0 + \beta_1 X_1 + \varepsilon$; $X_1 = \text{Uniform}(0, 100)$, $\beta_0 = 0$, $\beta_1 = 0.10$, ε is lognormal $(F_{\varepsilon}^{-1}(0.5) = 0, \sigma = 0.75)$, and $n = 150$. Each P value was estimated with $m + 1 = 10000$ permutations. Fine dotted 1 : 1 line is the expected cdf

centered on their appropriate distributional parameter, for example, $F_{\mathbf{e}}(\bar{x}|\mathbf{X_1}) \neq 0$, although $F_{\boldsymbol{\varepsilon}}(\mu|\mathbf{X_1}) = 0$. Instead, the estimated distributional parameter associated with 0 for $\mathbf{e_{red}}$ has random binomial sampling variation that needs to be incorporated into the permutation scheme to provide valid Type I error rates. A double permutation scheme has been proposed for providing valid Type I errors for these hypotheses [5, 8, 16]. The first step uses a random binomial variable, $\tau^* \sim \text{binomial}(\tau, n)$, to determine the value on which the residuals $\mathbf{e_{red}}$ are centered, $\mathbf{e_{red}}^* = \mathbf{e_{red}} - F_{\mathbf{e}}^{-1}(\boldsymbol{\tau}^*|\mathbf{X}_1)$, and the second step permutes the randomly centered residuals $\mathbf{e_{red}}^*$ to the matrix $\mathbf{X}$. For least squares regression, τ is taken as 0.5. An example of Type I error rates for the double permutation compared to the uncorrected standard permutation test for the hypothesis $H_0 : \beta_0 = 0$ for a 0.5 quantile regression model is shown in Figure 3.

Example Application

In applications, we often make use of the fact that confidence intervals on parameters in a linear model may be constructed by inversion of permutation tests [5, 7, 18]. We obtain a $(1 - \alpha) \times 100\%$ confidence interval on a single parameter for a variable $\mathbf{x}_2 = \mathbf{X}_2$ by making the transformation $\mathbf{y} - \mathbf{x}_2\xi$ on a sequence of values of ξ and collecting those values that have $P \geq \alpha$ for a test of the null hypothesis $H_0 : \beta_2 = \xi$.

The data in Figure 1 were from a study designed to evaluate changes in Lahontan cutthroat trout (*Oncorhyncus clarki henshawi*) densities as a function of stream channel morphology as it varies over the semidesert Lahontan basin of northern Nevada, USA [10]. The quantile regression analyses published in [10] used inferential procedures based on asymptotic distributional evaluations of the τ-quantile rank score statistic. Here, for a selected subset of quantiles, the 90% confidence intervals and hypothesis of zero slope were made with permutation tests based on the T statistic and permuting residuals $\mathbf{e_{red}}$. Because

the null model for the weighted estimates was implicitly forced through the origin, the double permutation scheme was required to provide valid Type I error rates. It is obvious in this location-scale model that restricting estimation and inferences to central distributional parameters, whether the mean or median, would have failed to detect changes in the trout densities at lower and higher portions of the distribution. Note that the permutation-based 90% confidence intervals for the weighted least squares estimate do not differ appreciably from the usual F distribution evaluation of the statistic nor do the intervals for the quantile estimates differ much from those based on the quantile rank score inversion [6].

Software

Computer routines for performing permutation tests on linear, least squares regression models are available in the Blossom software available from the U. S. Geological Survey (`www.fort.usgs.gov/products/software/sofware.asp`), in RT available from Western EcoSystems Technology, Inc.(`www.west-inc.com/`), in NPMANOVA from the web page of M. J. Anderson (`www.stat.auckland.ac.nz/~mja/`), and from the web page of P. Legendre (`www.fas.umontreal.ca/BIOL/legendre/`). The Blossom software package also features permutation tests for median and quantile regression. General permutation routines that can be customized for testing linear models are available in S-Plus and R.

References

[1] Anderson, M.J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression, *Canadian Journal of Fisheries and Aquatic Science* **58**, 626–639.

[2] Anderson, M.J. & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model, *Journal of Statistical Computation and Simulation* **62**, 271–303.

[3] Anderson, M.J. & Robinson, J. (2001). Permutation tests for linear models, *Australian New Zealand Journal of Statistics* **43**, 75–88.

[4] Anderson, M.J. & TerBraak, C.J.F. (2003). Permutation tests for multi-factorial analysis of variance, *Journal of Statistical Computation and Simulation* **73**, 85–113.

[5] Cade, B.S. (2003). *Quantile regression models of animal habitat relationships*, Ph.D. dissertation, Colorado State University, Fort Collins.

[6] Cade, B.S. & Noon, B.R. (2003). A gentle introduction to quantile regression for ecologists, *Frontiers in Ecology and the Environment* **1**, 412–420.

[7] Cade, B.S. & Richards, J.D. (1996). Permutation tests for least absolute deviation regression, *Biometrics* **52**, 886–902.

[8] Cade, B.S. & Richards, J.D. (In review). A drop in dispersion permutation test for regression quantile estimates. *Journal of Agricultural, Biological, and Environmental Statistics*.

[9] Commenges, D. (2003). Transformations which preserve exchangeability and application to permutation tests, *Journal of Nonparametric Statistics* **15**, 171–185.

[10] Dunham, J.B., Cade, B.S. & Terrell, J.W. (2002). Influences of spatial and temporal variation on fish-habitat relationships defined by regression quantiles, *Transactions of the American Fisheries Society* **131**, 86–98.

[11] Freedman, D. & Lane, D. (1983). A nonstochastic interpretation of reported significance levels, *Journal of Business and Economic Statistics* **1**, 292–298.

[12] Good, P. (2002). Extensions of the concept of exchangeability and their applications, *Journal of Modern Applied Statistical Methods* **1**, 243–247.

[13] Kennedy, P.E. & Cade, B.S. (1996). Randomization tests for multiple regression, *Communications in Statistics – Simulation and Computation* **25**, 923–936.

[14] Koenker, R. & Bassett, G. (1978). Regression quantiles, *Econometrica* **46**, 33–50.

[15] Koenker, R. & Hallock, K.F. (2001). Quantile regression, *Journal of Economic Perspectives* **15**,(4), 143–156.

[16] Legendre, P. & Desdevises, Y. (2002). *Independent Contrasts and Regression Through the Origin*, (In review, available at `www.fas.umontreal.ca/BIOL/legendre/reprints/index.html`).

[17] LePage, R. & Podgórski, K. (1993). Resampling permutations in regression with exchangeable errors, *Computing Science and Statistics* **24**, 546–553.

[18] Manly, B.F.J. (1997). *Randomization, Bootstrap, and Monte Carlo Methods in Biology*, 2nd Edition, Chapman & Hall, New York.

[19] Mielke Jr, P.W. & Berry, K.J. (2001). *Permutation Methods: A Distance Function Approach*, Springer-Verlag, New York.

[20] Pesarin, F. (2001). *Multivariate Permutation Tests: With Applications in Biostatistics*, John Wiley & Sons, New York.

[21] ter Braak, C.J.F. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA, in *Bootstrapping and Related Techniques*, K.H. Jockel, G. Rothe & W. Sendler, eds, Springer-Verlag, New York.

BRIAN S. CADE